

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Reyes, Gregory R  
Yarbough, Patrice O  
Bradley, Daniel W  
Krawczynski, Krzysztof Z  
Tam, Albert  
Fry, Kirk E
- (ii) TITLE OF INVENTION: DNA Sequences of Enterically Transmitted  
Non-A/Non-B Hepatitis Viral Agent
- (iii) NUMBER OF SEQUENCES: 20
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Dehlinger & Associates
  - (B) STREET: 350 Cambridge Avenue, Suite 250
  - (C) CITY: Palo Alto
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 94306
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 09/128,275
  - (B) FILING DATE: 03-AUG-1998
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/279,823
  - (B) FILING DATE: 25-JUL-1994
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 07/681,078
  - (B) FILING DATE: 05-APR-1991
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 07/505,888
  - (B) FILING DATE: 05-APR-1990
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 07/420,921
  - (B) FILING DATE: 13-OCT-1989
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 07/367,486
  - (B) FILING DATE: 16-JUN-1989
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 07/336,672
  - (B) FILING DATE: 11-APR-1989
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 07/208,997
  - (B) FILING DATE: 17-JUN-1988

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Petitory, Joanne R.  
(B) REGISTRATION NUMBER: 42,995  
(C) REFERENCE/DOCKET NUMBER: 4600-0133.24

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (650) 324-0880  
(B) TELEFAX: (650) 324-0960

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1295 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 1.33 kb EcoRI insert of ET1.1,  
forward sequence

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 1..1293

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 2..1294

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 3..1295

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGACCTGTCC CTGTTGCAGC TGTTCTACCA CCCTGCCCCG AGCTCGAACA GGGCCTTCTC	60
TACCTGCCCC AGGAGCTCAC CACCTGTGAT AGTGTGTA CATTGAATT AACAGACATT	120
GTGCACTGCC GCATGGCCGC CCCGAGCCAG CGCAAGGCCG TGCTGTCCAC ACTCGTGGGC	180
CGCTACGGCG GTCGCACAAA GCTCTACAAT GCTTCCCACT CTGATGTTCC GCACTCTCTC	240
GCCCGTTTTA TCCCAGCAT TGGCCCGTA CAGGTACAA CTTGTGAATT GTACGAGCTA	300
GTGGAGGCCA TGGTCGAGAA GGGCCAGGAT GGCTCCGCCG TCCTTGAGCT TGATCTTTGC	360
AACCGTGACG TGTCCAGGAT CACCTTCTTC CAGAAAGATT GTAACAAGTT CACCACAGGT	420
GAGACCATTC CCCATGGTAA AGTGGCCAG GGCATCTCCG CCTGGAGCAA GACCTTCTGC	480
GCCCTCTTTG GGCCTTGGTT CCGGCTATT GAGAAGGCTA TTCTGGCCCT GCTCCCTCAG	540
GGTGTGTTTT ACGGTGATGC CTTTATGATC ACGGTCTTCT CGGCGGCTGT GGCCGCAGCA	600

AAGGCATCCA TGGTGTGTTGA GAATGACTTT TGTGAGTTTG ACTCCACCCA GAATAACTTT	660
TCTGTGGGTC TAGAGTGTGC TATTATGGAG GAGTGTGGGA TGCCGCAGTG GTCATCCGC	720
CTGTATCACC TTATAAGGTC TGCCTGGATC TTGCAGGCCC CGAAGGAGTC TCTGCGAGGG	780
TTTTGGAAGA AACACTCCGG TGAGCCCGGC ACTCTTCTAT GGAATACTGT CTGGAATATG	840
GCCGTTATTA CCCACTGTTA TGAATTCCGC GATTTTCAGG TGGCTGCCTT TAAAGGTGAT	900
GATTCGATAG TGCTTTGCAG TGAGTATCGT CAGAGTCCAG GAGCTGCTGT CCTGATCGCC	960
GGCTGTGGCT TGAAGTTGAA GGTAGATTTC CGCCCGATCG GTTTGTATGC AGGTGTTGTG	1020
GTGGCCCCCG GDCCTGGGCG GCTCCCTGAT GTTGTGGGCT TCGCCGGCCG GCTTACCGAG	1080
AAGAATTGGG GDCCTGGGCG TGAGCGGGCG GAGCAGCTCC GCCTGCTGT TAGTGATTTC	1140
CTCCGCAAGC TCACGAATGT AGCTCAGATG TGTGTGGATG TTGTTTCCCG TGTTTATGGG	1200
GTTTCCCCTG GACTCGTTCA TAACCTGATT GGCATGCTAC AGGCTGTTGC TGATGGCAAG	1260
GCACATTCA CTGAGTCAAT AAAACCAAGT CTCGA	1295

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 431 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Arg	Pro	Val	Pro	Val	Ala	Ala	Val	Leu	Pro	Pro	Cys	Pro	Glu	Leu	Glu
1				5					10					15	
Gln	Gly	Leu	Leu	Tyr	Leu	Pro	Gln	Glu	Leu	Thr	Thr	Cys	Asp	Ser	Val
		20						25					30		
Val	Thr	Phe	Glu	Leu	Thr	Asp	Ile	Val	His	Cys	Arg	Met	Ala	Ala	Pro
		35					40					45			
Ser	Gln	Arg	Lys	Ala	Val	Leu	Ser	Thr	Leu	Val	Gly	Arg	Tyr	Gly	Gly
		50				55					60				
Arg	Thr	Lys	Leu	Tyr	Asn	Ala	Ser	His	Ser	Asp	Val	Arg	Asp	Ser	Leu
		65			70					75					80
Ala	Arg	Phe	Ile	Pro	Ala	Ile	Gly	Pro	Val	Gln	Val	Thr	Thr	Cys	Glu
			85					90						95	
Leu	Tyr	Glu	Leu	Val	Glu	Ala	Met	Val	Glu	Lys	Gly	Gln	Asp	Gly	Ser
			100					105					110		
Ala	Val	Leu	Glu	Leu	Asp	Leu	Cys	Asn	Arg	Asp	Val	Ser	Arg	Ile	Thr
		115					120					125			
Phe	Phe	Gln	Lys	Asp	Cys	Asn	Lys	Phe	Thr	Thr	Gly	Glu	Thr	Ile	Ala
		130					135					140			

His Gly Lys Val Gly Gln Gly Ile Ser Ala Trp Ser Lys Thr Phe Cys  
 145 150 155 160  
 Ala Leu Phe Gly Pro Trp Phe Arg Ala Ile Glu Lys Ala Ile Leu Ala  
 165 170 175  
 Leu Leu Pro Gln Gly Val Phe Tyr Gly Asp Ala Phe Asp Asp Thr Val  
 180 185 190  
 Phe Ser Ala Ala Val Ala Ala Ala Lys Ala Ser Met Val Phe Glu Asn  
 195 200 205  
 Asp Phe Ser Glu Phe Asp Ser Thr Gln Asn Asn Phe Ser Leu Gly Leu  
 210 215 220  
 Glu Cys Ala Ile Met Glu Glu Cys Gly Met Pro Gln Trp Leu Ile Arg  
 225 230 235 240  
 Leu Tyr His Leu Ile Arg Ser Ala Trp Ile Leu Gln Ala Pro Lys Glu  
 245 250 255  
 Ser Leu Arg Gly Phe Trp Lys Lys His Ser Gly Glu Pro Gly Thr Leu  
 260 265 270  
 Leu Trp Asn Thr Val Trp Asn Met Ala Val Ile Thr His Cys Tyr Asp  
 275 280 285  
 Phe Arg Asp Phe Gln Val Ala Ala Phe Lys Gly Asp Asp Ser Ile Val  
 290 295 300  
 Leu Cys Ser Glu Tyr Arg Gln Ser Pro Gly Ala Ala Val Leu Ile Ala  
 305 310 315 320  
 Gly Cys Gly Leu Lys Leu Lys Val Asp Phe Arg Pro Ile Gly Leu Tyr  
 325 330 335  
 Ala Gly Val Val Val Ala Pro Gly Leu Gly Ala Leu Pro Asp Val Val  
 340 345 350  
 Arg Phe Ala Gly Arg Leu Thr Glu Lys Asn Trp Gly Pro Gly Pro Glu  
 355 360 365  
 Arg Ala Glu Gln Leu Arg Leu Ala Val Ser Asp Phe Leu Arg Lys Leu  
 370 375 380  
 Thr Asn Val Ala Gln Met Cys Val Asp Val Val Ser Arg Val Tyr Gly  
 385 390 395 400  
 Val Ser Pro Gly Leu Val His Asn Leu Ile Gly Met Leu Gln Ala Val  
 405 410 415  
 Ala Asp Gly Lys Ala His Phe Thr Glu Ser Val Lys Pro Val Leu  
 420 425 430

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(d) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: DNA  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (vi) ORIGINAL SOURCE:  
       (c) INDIVIDUAL ISOLATE: linker - top (5') sequence  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:  
 GGAATTCGCG GCGGCTCG

13

(2) INFORMATION FOR SEQ ID NO:4:  
 (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 20 base pairs  
       (B) TYPE: nucleic acid  
       (C) STRANDEDNESS: single  
       (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: DNA  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (vi) ORIGINAL SOURCE:  
       (c) INDIVIDUAL ISOLATE: linker - bottom (3') sequence  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:  
 CGAGCGGCCG CGAATTCCTT

20

(2) INFORMATION FOR SEQ ID NO:5:  
 (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 1295 base pairs  
       (B) TYPE: nucleic acid  
       (C) STRANDEDNESS: double  
       (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: DNA  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (vi) ORIGINAL SOURCE:  
       (c) INDIVIDUAL ISOLATE: 1.33 kb EcoRI insert of ET1.1,  
               reverse sequence

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:  
 TCGAGCACTG GTTTTACTGA CTCAGTGAAA TGTGCCTTGC CATCAGCAAC AGCCTGTAGC  
 ATGCCAATCA GGTATGAAC GAGTCCAGGG GAAACCCCAT AAACACGGGA AACAAATCC

60

120

ACACACATCT GAGCTACATT CCGTACGCTG CGGAGGAAAT CACTAACAGC GAGGCGGAGC	180
TGCTCGGCCC GGTGAGGGCC AGGGGCCCCAA TTCTTCTCGG TAAGCCGCGC GGGGAAGCGC	240
ADAACATCAG GGAGCGCGCC AAGGCGGGGG GCCACCACAA CACCTGCATA CAAACCGATC	300
GGGCGGAAAT CTACCTTCAA CTTCAAGCCA CAGCGGGCGA TCAGGACAGC AGCTCCTGGA	360
CTCTGACGAT ACTCACTGCA AAGCACTATC GAATCATCAC CTTTAAAGGC AGCCACCTGA	420
AAATCGCGGA AGTCATAACA GTGGGTAATA ACGGCCATAT TCAGACAGT ATTCCATAGA	480
AGAGTGCCCG GGTGACCGGA GTGTTTCTTC CAAAACCCTC GCAGAGACTC CTTGGGGGCC	540
TGCAAGATCC ACGCAGACCT TATAAGSTGA TACAGGCGGA TGAGGCACTG CGGATCCCA	600
CACTCCTCCA TAATAGCACA CTCTAGACCC AGAGAAAAGT TATTCTGGGT GGAGTCAAAC	660
TCAGAAAAGT CATTCTCAGA CACCATGGAT GCGTTTCTG CGGCACAGC CGCGAGAAAG	720
ACGGTGTGAT CAAAGGCATC ACCGTAAAAC ACACCTGAG GGAGCAGGGC CAGAATAGCC	780
TTCTCAATAG CGCGGAACCA AGGCCCAAAG AGGGCGCAGA AGGTCTTGCT CCAGGCGGAG	840
ATGCGCTGCG CCACTTTACC ATGGGCAATG GTCTCACCTG TGGTGAACCT GTTACAATCT	900
TTCTGGAAGA AGGTGATCCT GGACACGTCA CGGTTGCAAA GATCAAGCTC AAGGACGGCG	960
GAGCCATCCT GCGCTTCTC GACCATGGCC TCCACTAGCT CGTACAATTC ACAAGTTGTA	1020
ACCTGTACGG GGCCAATGGC CGGGATAAAA CGGGCGAGAG AGTCGCGAAC ATCAGAGTGG	1080
GAAGCATTGT AGAGCTTTGT GCGACCGCCG TAGCGGCCCA CGAGTGTGGA CAGCACGGCC	1140
TTGCGCTGGC TCGGGGCGGC CATGCGGCAG TGCACAATGT CTGTTAATTC AAATGTTACG	1200
ACACTATCAC AGGTGGTGAG CTCCTGGGGC AGGTAGAGAA GGCCCTGTTC GAGCTCGGGG	1260
CAGGGTGGTA GAACAGCTGC AACAGGGACA GGTCT	1295

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7195 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: HEV - Burma strain

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 28..5106

(ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 5147..7126

(ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 5106..5474

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGSCAGACCA CATATGTGGT CGATGCCATG GAGGCCCATC AGTTTATTAA GGCTCCTGGC	60
ATCACTACTG CTATTGAGCA GGCTGCTCTA GCAGCGGCCA ACTCTGCCCT GGCGAATGCT	120
GTGGTAGTAA GGCTTTTCT CTCTCACCAG CAGATTGAGA TCCCTATTAA CCTAATGCAA	180
CCTCGCCAGC TTGTTTTTCG CCGCGAGGTT TTCTGGAATC ATCCCATCCA GCGTGTGATC	240
CATAACGAGC TGGAGCTTTA CTGCGCGCGC GCGTCCGCGC GGTGTCTTGA AATTGGCGCC	300
CATCCCGGCT CAATAAATGA TAATCCTAAT GTGGTCCACC GGTGTCTCCT CGGCGCTGTT	360
GGCGGTGATG TTCAGCGCTG GTATACTGCT CCGACTCGCG GGCGGGCTGC TAATTGCGGG	420
CCTTCGGCGC TCGCGCGGCT TCGCGCTGCT GACCGCACTT ACTGGCTCGA CCGGTTTTCT	480
GGCTGTAACT TCGCGCGCGA GACTGCGATC GCGCTCTACT CCTTCATGA TATGTCACCA	540
TCTGATGTGG CCGAGGCCAT GTTCGGCCAT GGTATGACGC GGCTCTATGC CGCGCTCCAT	600
CTTCGCGCTG AGGTCTGCT GCCCGCTGGC ACATATCGCA CCGCATCGTA TTTGCTAATT	660
CATGACGGTA GCGCGGTTGT GGTGACGTAT GAGGGTGATA CTAGTGTGG TTACAACCAC	720
GATGTCTCCA ACTGCGCTC CTGGATTAGA ACCACCAAGG TTACCGGAGA CCATCCCCCTC	780
GTTATCGAGC GGGTTAGGGC CATTGGCTGC CACTTTGTTC TCTTGCTCAC GGCAGCCCCG	840
GAGCCATCAC CTATGCCTTA TGTTCCTTAC CCGCGGTCTA CCGAGGTCTA TGTCCGATCG	900
ATCTTCGGCC CGGGTGGCAC CCCTTCCTTA TTCCCAACCT CATGCTCCAC TAAGTCGACC	960
TTCCATGCTG TCCCTGCCCA TATTTGGGAC CGTCTTATGC TGTTCGGGGC CACCTTGGAT	1020
GACCAAGCCT TTTGCTGCTC CCGTTTAATG AACTACCTTC GCGGCATTAG CTACAAGGTC	1080
ACTGTTGGTA CCCTTGTTGC TAATGAAGGC TGAATGCCT CTGAGGACGC CCTCACAGCT	1140
GTTATCACTG CCGCTACCT TACCAATTGC CACGAGCGGT ATCTCCGCAC CAGGCTATA	1200
TCCAAGGGGA TCGTCTGCTT GGAAGGGGAG CATGCCGAGA AGTTTATAAC AGGCTCTAC	1260
AGTGGCTCT TCGAGAAGTC CGGCGGTGAT TACATCCCTG GCGTCACTT GGAGTTCTAC	1320
GCCCACTGCA GCGGCTGGCT CTCGCGCGGC TTTCATCTTG ATCCACGGGT GTTGGTTTTT	1380
GAGGASTGCG CCGGCTGCGA TTGTAGGACC GCGATCGTA AGGCGCTCTC AAAGTTTTGC	1440
TCTTCATGA AGTGGCTTGC TCAGGASTGC AACTGCTTCC TTCAGGCTGC AGAAGCGCGC	1500
GTGCGCGACG AGGTTCATGA TAATGAAGCC TATGAGGGGT CCGATGTTGA CCTGCTGAG	1560

TCGCGCATTG GTGACATATC TGGGTGCTAT GTGCTGCTG GCACTGCTGT GCAACCGCTC	1620
TACGAGGCTG TCGATCTGCG GCGTGAGATT GTGGGTGCTG GGGGCTGCTT GACGCGCACA	1680
GTAAAGGTCT CCCAGGTGCA TGGGCGGATC GATTGGGAGA CCGTTCTTGG TAACAAAACC	1740
TTTCGCACGT CGTTCGTTGA GGGGGGGGTC TTAGAGACCA ATGGCCGACA GCGGCACAAAT	1800
CTCTCCTTCG ATGCCAGTCA GAGCACTATG GCGGCTGGCC CTTTCAGTCT CAGGTATGCC	1860
GCCTCTGCAG CTGGGCTGGA GGTGGGCTAT GTTGTGCGCG GGCTTGACCA TCGGGCGGTT	1920
TTTGCCGCGG GTGTTTCACC CCGGTCAGCC GCGGGGAGG TTACGGCTTT CTGCTCTGCC	1980
CTATACAGGT TTAACCGTGA GGGGAGCGC GATTGGGTGA TCGGTAACTT ATGGTTCGAT	2040
CCTGAGGGAC TCATTGGCTT CTGGGCGCGG TTTTGCGCGG GCGATGTTG GGAATCGGCT	2100
AATGCATTCT GTGGGAGAG CACACTTTAC ACCGCTACTT GGTGGAGGT TGATGCGCTC	2160
TCTAGTCCAG CCGGGGCTGA CTTAGGTTTT ATGTGTGAGC CTTCTATACC TAGTAGGGCC	2220
GCCAGGCTTA GCGTGGGCGC GCGGTACCC CCGGCTGAC CCGAGCGCTT CCGTCTCTCC	2280
TCTGCGCGG GCGTGGGTGA GCGGCTTCT GCGGCTACCG CCGGGGGCGC GCGCTAACT	2340
CACGAGACGG CCGGGCACCG GCGGCTGCTC TTCAGCTACC CCGATGGCTC TAAGGTATTC	2400
GCGGCTCGC TGTTCGAGTC GACATGCACG TGGCTCGTTA ACGCGTCTAA TGTTGACCAC	2460
CGCCCTGGCG GCGGGCTTTG CCATGCATTT TACCAAAGGT ACCCGGCTC CTTTGATGCT	2520
GCCTCTTTTG TGATGCGCGA CCGCGGCGCC GCGTACACAC TAACCGCGCG GCCAATAATT	2580
CACGCTGTG CCCCTGATTA TAGGTTGGAA CATAACCCAA AGAGGCTTGA GGCTGCTTAT	2640
CGGGAAACTT GCTCCCGCCT CCGCACCGCT GCATACCGC TCCTCGGGAC CGGCATATAC	2700
CAGGTGCCGA TCGGCCCCAG TTTTGACGCC TGGGAGCGGA ACCACCGCCC CCGGGATGAG	2760
TTGTACCTTC CTGAGCTTGC TGCCAGATGG TTTGAGGCCA ATAGGCCGAC CCGCCCCACT	2820
CTCACTATAA CTGAGGATGT TGCACGGACA GCGAATCTGG CCATCGAGCT TGAATCAGCC	2880
ACAGATGTG GCCGGGCTG TGCGGCTGT CCGGTACCC CCGGCGTTGT TCAGTACCAG	2940
TTTACTGCAG GTGTGCTGG ATCGGCAAG TCGGCTCTA TCACCGAAGC CGATGTGGAC	3000
GTTGTCTGG TCGGAGCGG TGASTTGGCT AATGCTTGG GCGGTGCGG CTTTGTGCT	3060
TTTACCGCGC ATACTGCGCG CAGAGTCACC CAGGGCGCGT GGGTGTGCT TGAATAGGCT	3120
CCATGCTCT CCGCTCAGCT GGTGTGCTC CAGATGCAGC GCGCGGCGAC CGTCCAGCTT	3180
CTTGCGGACC CGAACCAGAT CCGAGGCTG GACTTTGAGC ACGCTGGGCT CGTCTCGGCG	3240
ATCAGGCGCG AATTAGGCGC CAGTCTCTGG TGCGATCTTA CCGATCGCTG GCTGTGGAT	3300
GTATGCGAGC TCATCGCTGG TGCATACCC ATGATCGAGA CCACTAGCGG GGTCTCTCGT	3360
TGCTTGTCTT GCGGTGAGCG TGCGGTGCGG CAGAACTAG TGTTCACCGA GCGGGCGAAG	3420



CCCTCCCAAC	CTGGCTCACT	CAGGCTCCAC	CAGGCTCCAC	CCCTCACTCA	CAGGCTCCAC	3480
ACTATTATTG	CCACAGCAGA	TCCCGCGGGC	CTTATTCACT	CGTCTCGGGC	TCATGCCATT	3540
GTTGCTCTGA	CGCGCCACAC	TGAGAACTGC	GTCATCATTG	ACGCACGAGG	CTGCTTGGC	3600
SAGGTGGCA	TCTCCSATGC	AATCGTTAAT	AACTTTTCC	TGCTGGTGG	CGAAATTGTT	3660
CACGAGCTGC	CATCACTTAT	TCCCGCTGC	AACTCTGAGG	CCAATGTTGA	CACCTCTGCT	3720
GCCTTCCCGC	TGTCTTGCCA	GATTAGTGCC	TTCCATCACT	TGGCTGAGGA	GTTTGGGCAC	3780
AGACCTGTGC	CTGTTGAGC	TGTTCTACCA	CCCTGCGCGG	AGCTCGAACA	GGGCTTCTGC	3840
TACTTCCCGC	AGGAGCTCAC	CACCTGTGAT	AGTGTCTTAA	CATTTGAATT	AACAGACATT	3900
GTCACTTGC	GCATGGCGGC	CCCGAGCGAG	CGCAAGCGCG	TGCTCTCCAC	ACTCTTGGGC	3960
CCCTAGCGCG	GTGCGACAAA	GCTCTACAAT	GGTCTCCACT	CTGATTTGG	CGATCTCTGC	4020
GGCGGTTTTA	TCCCGCGCAT	TGCGCGCGTA	CAGGTTACAA	CTTGTGAATT	GTAGGAGCTA	4080
GTGGAGGCGA	TGCTCGAGAA	GGCGAGGAT	GGCTCGCGCG	TGCTTGAGCT	TGATCTTTGC	4140
AACCGTGAGC	TGCTCAGGAT	CACCTTCTTC	CAGAAAGATT	GTAACAGATT	CACGACAGGT	4200
GAGACATTG	CCCATGCTAA	AGTGGCGCAG	GGCATCTCGG	CGTGGAGCAA	GACCTTCTGC	4260
GCCCTCTTTG	GGCCTTGGTT	CCGCGCTATT	GAGAAGGCTA	TTCTCGCCCT	GCTCCCTCAG	4320
GGTGTGTTTT	ACGGTGATGC	CTTTGATGAC	ACCGTCTTCT	CGCGCGCTGT	GGCGCAGCA	4380
AAGGCATCCA	TGGTGTGTTA	GAATGACTTT	TCTGAGTTTG	ACTCCACCCA	GAATACTTTT	4440
TCTCTGGGTC	TAGAGTGTGC	TATTATGGAG	GAGTGTGGGA	TGCGCGAGTG	GCTCATCCGC	4500
CTGTATCACC	TTATAAGGTC	TGCGTGGATC	TTGCAGGCCC	CGAAGGAGTC	TCTGCGAGGG	4560
TTTTGGAAGA	AACACTCCGG	TGAGCCCGGC	ACTCTTCTAT	GGAATACTGT	CTGGAATATG	4620
GCCGTTATTA	CCCACTGTTA	TGACTTCCGC	GATTTTCAGG	TGGCTGCCTT	TAAAGGTGAT	4680
GATTGATAG	TGCTTTGCAG	TGAGTATCGT	CAGAGTCCAG	GAGCTGCTGT	CCTGATCGCC	4740
GGCTGTGGCT	TGAAGTTGAA	GGTAGATTTG	CGCGCGATCG	GTTTGTATGC	AGGTGTTGTC	4800
GTGGCCCCCG	GCCTTGGCGC	GCTCCCTGAT	GTTGTGCGCT	TGCGCGCGCG	GCTTACCGAG	4860
AAGAAATGGG	GCCTTGGCGC	TGAGCGCGCG	GAGCAGCTCG	GCCTCGCTGT	TAGTGATTTC	4920
CTCGCAAGC	TCAAGAAATG	AGCTCAGATG	TGTGTGATG	TTGTTTCCCG	TGTTTATGGG	4980
GTTTCCCGCT	GACTCGTTCA	TAACTGATT	GGCATGCTAC	AGCTGTGTC	TGATGGCAAG	5040
GCATATTTCA	CTGAGTCACT	AAAACGAGTC	CTCGACTTGA	CAAAATCAAT	CTTGTGTGCG	5100
GTGGAATGAA	TAACATGTCT	TTTGTGCGCG	CGATGGGTTG	GGGACCATGC	GGCCTCGCGC	5160
TATTTTGTTC	CTGCTCTCTA	TGTTTTTGCC	TATGCTGCGC	GGGCGACCGC	CGGCTCAGCG	5220
GTCTGGCGCG	CGTCTGCGCG	GGCGCAGCGG	CGGTTCCCGC	GTTGGTTTCT	GGGCTGACCG	5280

AGTGGATTCT GAGCGCTTGG GAATGCGCTA TATTGATGCA AGGAAGCGCT TCGCGCGCGA	5340
TGTGACCGCT GGTGGCGCGG GTGACGCTGG TGTTGCGGAA CGCGCGCGAG CACTCGGCTG	5400
CGCTTGGGCT GACCAGGCGG AGCGCGCGCG CGTTGCGCTA CGTGCTAGAC CTACCAAGC	5460
TGGCGCGCGG CGGCTAAGCG CGGTGCGCTG GGCCCATGAC AGCGCGCGAG TGGCTGATGT	5520
GAATGCGCGG GGTGGCATCT TGCGCGCGGA GTATAAGCTA TGAACATCTC CGCTTACCTC	5580
TTCTGTGGCG ACCGGCACTA ACCTGGTTCT TTATGCGCGC CGTCTTAGTC CGCTTTTACC	5640
CGTTGAGGAC GGCACCAATA CGCATATAAT GGCACCGGAA GTTCTAATT ATGCGGASTA	5700
CGTGGTTGCG CGTGGCGGAA TCGTTAGCGG CGTGGTGGTC CGCAATGCTG TCGCGGTTA	5760
CGCATCTCTC ATCTCATCTC GGCACAGAC CACCAAGCG CGGAGCTCGG TTGATATGAA	5820
TTCAATAAGC TGGACGGATG TTCTATTTT AGTGGCGCGG GGCATAGCGT CTGAGCTTGT	5880
GATCGCAAGT GAGCGGCTAC ACTATCGTAA GCAAGGCTGG CGCTCGGTCG AGACTCTCTG	5940
GGTGGTTGAG GAGGAGGCTA CGTCTGGTCT TGTTATGCTT TGCATACATG GCTCACTCT	6000
AAATTCTAT ACTAATACAC CGTATACCGG TGCGCTCGGG CTGTTGGACT TTGCCCTTGA	6060
GCTTGAGTTT CGCAACCTTA CGCGCGGTAA CACCAATAG CGGGTCTCCC GTTATTCGAG	6120
CACTGCTCGC CACCGCCTTC GTCGCGGTGC GGACGGGACT GCCGAGCTCA CCACCACGGC	6180
TGCTACCGCG TTTATGAAGG ACCTCTATTT TACTAGTACT AATGGTGTCT GTGAGATCGG	6240
CGCGCGGATA GCCCTCACCC TGTCAACCT TGCTGACACT CTGCTTGGCG GCCTGCCGAC	6300
AGAATTGATT TCGTCGGCTG GTGGCGAGCT GTTCTACTCC CGTCCCGTTG TCTCAGCCAA	6360
TGGCGAGCCG ACTGTTAAGT TGTATACATC TGTAGAGAAT GCTCAGCAGG ATAAGGGTAT	6420
TGCAATCCCG CATGACATTG ACCTCGGAGA ATCTCGTGTG GTTATTCAGG ATTATGATAA	6480
CCAACATGAA CAAGATCGGC CGACGCCTTC TCAGCCCCA TCGCGCCCTT TCTCTGTCT	6540
TCAGCTAAT GATGTGCTTT GGCTCTCTCT CACCGCTGCG GAGTATGACC AGTCCACTTA	6600
TGGCTCTTCG ACTGGCCCAG TTTATGTTTC TGACTCTGTG ACCTTGGTTA ATGTTGGGAC	6660
CGCGCGCGAG GCCGTTGCCC GGTGCTCGA TTGGACCAAG GTCACACTTG ACGGTCGCGT	6720
CGTCTCGAGC ATCGAGAGT ACTGGAAGAG CTCTTTTCTC CTGCGGCTCG GCGGTAAGCT	6780
CTCTTTCTGG GAGGCGAGCA CAACTAAAGC CGGTACCTT TATAATTATA ACACCACTGC	6840
TAGCGACCAA CTGCTTGTCT AGAATGCGCG CGGGCACCGG GTCGCTATTT CCACTTACAC	6900
CACTAGCTTG GGTGCTGGTC CGGTCTCAT TTCTGGGTT GCGCTTTTAG CGCGGCACTC	6960
TGGCTAGCA TTGCTTGAGG ATACCTTGA CTACCTGCG CGCGCGGATA CTTTGTATGA	7020
TTTCTGCGCA GAGTGGCGCG CCCTTGGGCT TCAGGCGTGG GCTTTGAGT CTACTGTGCG	7080
TGAGCTTCAG CGCCTTAAGA TGAAGTGGG TAAAGCTCG GAGTTGTAGT TTATTTGCTT	7140

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1693 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

Met Glu Ala His Gln Phe Ile Lys Ala Pro Gly Ile Thr Thr Ala Ile
 1           5           10           15
Glu Gln Ala Ala Leu Ala Ala Ala Asn Ser Ala Leu Ala Asn Ala Val
          20           25           30
Val Val Arg Pro Phe Leu Ser His Gln Gln Ile Glu Ile Leu Ile Asn
          35           40           45
Leu Met Gln Pro Arg Gln Leu Val Phe Arg Pro Glu Val Phe Trp Asn
          50           55           60
His Pro Ile Gln Arg Val Ile His Asn Glu Leu Glu Leu Tyr Cys Arg
          65           70           75           80
Ala Arg Ser Gly Arg Cys Leu Glu Ile Gly Ala His Pro Arg Ser Ile
          85           90           95
Asn Asp Asn Pro Asn Val Val His Arg Cys Phe Leu Arg Pro Val Gly
          100          105          110
Arg Asp Val Gln Arg Trp Tyr Thr Ala Pro Thr Arg Gly Pro Ala Ala
          115          120          125
Asn Cys Arg Arg Ser Ala Leu Arg Gly Leu Pro Ala Ala Asp Arg Thr
          130          135          140
Tyr Cys Leu Asp Gly Phe Ser Gly Cys Asn Phe Pro Ala Glu Thr Gly
          145          150          155          160
Ile Ala Leu Tyr Ser Leu His Asp Met Ser Pro Ser Asp Val Ala Glu
          165          170          175
Ala Met Phe Arg His Gly Met Thr Arg Leu Tyr Ala Ala Leu His Leu
          180          185          190
Pro Pro Glu Val Leu Leu Pro Pro Gly Thr Tyr Arg Thr Ala Ser Tyr
          195          200          205
Leu Leu Ile His Asp Gly Arg Arg Val Val Val Thr Tyr Glu Gly Asp
          210          215          220
Thr Ser Ala Gly Tyr Asn His Asp Val Ser Asn Leu Arg Ser Trp Ile
          225          230          235          240
Arg Thr Thr Lys Val Thr Gly Asp His Pro Leu Val Ile Glu Arg Val

```

245

250

255

Arg Ala Ile Gly Cys His Phe Val Leu Leu Leu Thr Ala Ala Pro Glu  
 260 265 270  
 Pro Ser Pro Met Pro Tyr Val Pro Tyr Pro Arg Ser Thr Glu Val Tyr  
 275 280 285  
 Val Arg Ser Ile Phe Gly Pro Gly Gly Thr Pro Ser Leu Phe Pro Thr  
 290 295 300  
 Ser Cys Ser Thr Lys Ser Thr Phe His Ala Val Pro Ala His Ile Trp  
 305 310 315 320  
 Asp Arg Leu Met Leu Phe Gly Ala Thr Leu Asp Asp Gln Ala Phe Cys  
 325 330 335  
 Cys Ser Arg Leu Met Thr Tyr Leu Arg Gly Ile Ser Tyr Lys Val Thr  
 340 345 350  
 Val Gly Thr Leu Val Ala Asn Glu Gly Trp Asn Ala Ser Glu Asp Ala  
 355 360 365  
 Leu Thr Ala Val Ile Thr Ala Ala Tyr Leu Thr Ile Cys His Gln Arg  
 370 375 380  
 Tyr Leu Arg Thr Gln Ala Ile Ser Lys Gly Met Arg Arg Leu Glu Arg  
 385 390 395 400  
 Glu His Ala Gln Lys Phe Ile Thr Arg Leu Tyr Ser Trp Leu Phe Glu  
 405 410 415  
 Lys Ser Gly Arg Asp Tyr Ile Pro Gly Arg Gln Leu Glu Phe Tyr Ala  
 420 425 430  
 Gln Cys Arg Arg Trp Leu Ser Ala Gly Phe His Leu Asp Pro Arg Val  
 435 440 445  
 Leu Val Phe Asp Glu Ser Ala Pro Cys His Cys Arg Thr Ala Ile Arg  
 450 455 460  
 Lys Ala Leu Ser Lys Phe Cys Cys Phe Met Lys Trp Leu Gly Gln Glu  
 465 470 475 480  
 Cys Thr Cys Phe Leu Gln Pro Ala Glu Gly Ala Val Gly Asp Gln Gly  
 485 490 495  
 His Asp Asn Glu Ala Tyr Glu Gly Ser Asp Val Asp Pro Ala Glu Ser  
 500 505 510  
 Ala Ile Ser Asp Ile Ser Gly Ser Tyr Val Val Pro Gly Thr Ala Leu  
 515 520 525  
 Gln Pro Leu Tyr Gln Ala Leu Asp Leu Pro Ala Glu Ile Val Ala Arg  
 530 535 540  
 Ala Gly Arg Leu Thr Ala Thr Val Lys Val Ser Gln Val Asp Gly Arg  
 545 550 555 560  
 Ile Asp Cys Glu Thr Leu Leu Gly Asn Lys Thr Phe Arg Thr Ser Phe  
 565 570 575

Val Asp Gly Ala Val Leu Glu Thr Asn Gly Pro Glu Arg His Asn Leu  
 590 595 590  
 Ser Phe Asp Ala Ser Gln Ser Thr Met Ala Ala Gly Pro Phe Ser Leu  
 595 600 605  
 Thr Tyr Ala Ala Ser Ala Ala Gly Leu Glu Val Arg Tyr Val Ala Ala  
 610 615 620  
 Gly Leu Asp His Arg Ala Val Phe Ala Pro Gly Val Ser Pro Arg Ser  
 625 630 635 640  
 Ala Pro Gly Glu Val Thr Ala Phe Cys Ser Ala Leu Tyr Arg Phe Asn  
 645 650 655  
 Arg Glu Ala Gln Arg His Ser Leu Ile Gly Asn Leu Trp Phe His Pro  
 660 665 670  
 Glu Gly Leu Ile Gly Leu Phe Ala Pro Phe Ser Pro Gly His Val Trp  
 675 680 685  
 Glu Ser Ala Asn Pro Phe Cys Gly Glu Ser Thr Leu Tyr Thr Arg Thr  
 690 695 700  
 Trp Ser Glu Val Asp Ala Val Ser Ser Pro Ala Arg Pro Asp Leu Gly  
 705 710 715 720  
 Phe Met Ser Glu Pro Ser Ile Pro Ser Arg Ala Ala Thr Pro Thr Leu  
 725 730 735  
 Ala Ala Pro Leu Pro Pro Pro Ala Pro Asp Pro Ser Pro Pro Ser  
 740 745 750  
 Ala Pro Ala Leu Ala Glu Pro Ala Ser Gly Ala Thr Ala Gly Ala Pro  
 755 760 765  
 Ala Ile Thr His Gln Thr Ala Arg His Arg Arg Leu Phe Thr Tyr  
 770 775 780  
 Pro Asp Gly Ser Lys Val Phe Ala Gly Ser Leu Phe Glu Ser Thr Cys  
 785 790 795 800  
 Thr Trp Leu Val Asn Ala Ser Asn Val Asp His Arg Pro Gly Gly Gly  
 805 810 815  
 Leu Cys His Ala Phe Tyr Gln Arg Tyr Pro Ala Ser Phe Asp Ala Ala  
 820 825 830  
 Ser Phe Val Met Arg Asp Gly Ala Ala Ala Tyr Thr Leu Thr Pro Arg  
 835 840 845  
 Pro Ile Ile His Ala Val Ala Pro Asp Tyr Arg Leu Glu His Asn Pro  
 850 855 860  
 Lys Arg Leu Glu Ala Ala Tyr Arg Glu Thr Cys Ser Arg Leu Gly Thr  
 865 870 875 880  
 Ala Ala Tyr Pro Leu Leu Gly Thr Gly Ile Tyr Gln Val Pro Ile Gly  
 885 890 895  
 Pro Ser Phe Asp Ala Trp Glu Arg Asn His Arg Pro Gly Asp Glu Leu  
 900 905 910

Tyr Leu Pro Glu Leu Ala Ala Arg Trp Phe Glu Ala Asn Arg Pro Thr  
 915 925 925  
 Arg Pro Thr Leu Thr Ile Thr Glu Asp Val Ala Arg Thr Ala Asn Leu  
 930 935 940  
 Ala Ile Glu Leu Asp Ser Ala Thr Asp Val Gly Arg Ala Cys Ala Gly  
 945 950 955 960  
 Cys Arg Val Thr Pro Gly Val Val Gln Tyr Gln Phe Thr Ala Gly Val  
 965 970 975  
 Pro Gly Ser Gly Lys Ser Arg Ser Ile Thr Gln Ala Asp Val Asp Val  
 980 985 990  
 Val Val Val Pro Thr Arg Glu Leu Arg Asn Ala Trp Arg Arg Arg Gly  
 995 1000 1005  
 Phe Ala Ala Phe Thr Pro His Thr Ala Ala Arg Val Thr Gln Gly Arg  
 1010 1015 1020  
 Arg Val Val Ile Asp Glu Ala Pro Ser Leu Pro Pro His Leu Leu Leu  
 1025 1030 1035 1040  
 Leu His Met Gln Arg Ala Ala Thr Val His Leu Leu Gly Asp Pro Asn  
 1045 1050 1055  
 Gln Ile Pro Ala Ile Asp Phe Glu His Ala Gly Leu Val Pro Ala Ile  
 1060 1065 1070  
 Arg Pro Asp Leu Gly Pro Thr Ser Trp Trp His Val Thr His Arg Trp  
 1075 1080 1085  
 Pro Ala Asp Val Cys Glu Leu Ile Arg Gly Ala Tyr Pro Met Ile Gln  
 1090 1095 1100  
 Thr Thr Ser Arg Val Leu Arg Ser Leu Phe Trp Gly Glu Pro Ala Val  
 1105 1110 1115 1120  
 Gly Gln Lys Leu Val Phe Thr Gln Ala Ala Lys Pro Ala Asn Pro Gly  
 1125 1130 1135  
 Ser Val Thr Val His Glu Ala Gln Gly Ala Thr Tyr Thr Glu Thr Thr  
 1140 1145 1150  
 Ile Ile Ala Thr Ala Asp Ala Arg Gly Leu Ile Gln Ser Ser Arg Ala  
 1155 1160 1165  
 His Ala Ile Val Ala Leu Thr Arg His Thr Glu Lys Cys Val Ile Ile  
 1170 1175 1180  
 Asp Ala Pro Gly Leu Leu Arg Glu Val Gly Ile Ser Asp Ala Ile Val  
 1185 1190 1195 1200  
 Asn Asn Phe Phe Leu Ala Gly Gly Glu Ile Gly His Gln Arg Pro Ser  
 1205 1210 1215  
 Val Ile Pro Arg Gly Asn Pro Asp Ala Asn Val Asp Thr Leu Ala Ala  
 1220 1225 1230  
 Phe Pro Pro Ser Cys Gln Ile Ser Ala Phe His Gln Leu Ala Glu Glu

1235

1240

1245

Leu Gly His Arg Pro Val Pro Val Ala Ala Val Leu Pro Pro Cys Pro  
 1250 1255 1260

Glu Leu Glu Gln Gly Leu Leu Tyr Leu Pro Gln Glu Leu Thr Thr Cys  
 1265 1270 1275 1280

Asp Ser Val Val Thr Phe Glu Leu Thr Asp Ile Val His Cys Arg Met  
 1285 1290 1295

Ala Ala Pro Ser Gln Arg Lys Ala Val Leu Ser Thr Leu Val Gly Arg  
 1300 1305 1310

Tyr Gly Gly Arg Thr Lys Leu Tyr Asn Ala Ser His Ser Asp Val Arg  
 1315 1320 1325

Asp Ser Leu Ala Arg Phe Ile Pro Ala Ile Gly Pro Val Gln Val Thr  
 1330 1335 1340

Thr Cys Glu Leu Tyr Glu Leu Val Glu Ala Met Val Glu Lys Gly Gln  
 1345 1350 1355 1360

Asp Gly Ser Ala Val Leu Glu Leu Asp Leu Cys Asn Arg Asp Val Ser  
 1365 1370 1375

Arg Ile Thr Phe Phe Gln Lys Asp Cys Asn Lys Phe Thr Thr Gly Glu  
 1380 1385 1390

Thr Ile Ala His Gly Lys Val Gly Gln Gly Ile Ser Ala Trp Ser Lys  
 1395 1400 1405

Thr Phe Cys Ala Leu Phe Gly Pro Trp Phe Arg Ala Ile Glu Lys Ala  
 1410 1415 1420

Ile Leu Ala Leu Leu Pro Gln Gly Val Phe Tyr Gly Asp Ala Phe Asp  
 1425 1430 1435 1440

Asp Thr Val Phe Ser Ala Ala Val Ala Ala Lys Ala Ser Met Val  
 1445 1450 1455

Phe Glu Asn Asp Phe Ser Glu Phe Asp Ser Thr Gln Asn Asn Phe Ser  
 1460 1465 1470

Leu Gly Leu Glu Cys Ala Ile Met Glu Glu Cys Gly Met Pro Gln Trp  
 1475 1480 1485

Leu Ile Arg Leu Tyr His Leu Ile Arg Ser Ala Trp Ile Leu Gln Ala  
 1490 1495 1500

Pro Lys Glu Ser Leu Arg Gly Phe Trp Lys Lys His Ser Gly Glu Pro  
 1505 1510 1515 1520

Gly Thr Leu Leu Trp Asn Thr Val Trp Asn Met Ala Val Ile Thr His  
 1525 1530 1535

Cys Tyr Asp Phe Arg Asp Phe Gln Val Ala Ala Phe Lys Gly Asp Asp  
 1540 1545 1550

Ser Ile Val Leu Cys Ser Glu Tyr Arg Gln Ser Pro Gly Ala Ala Val  
 1555 1560 1565

Leu Ile Ala Gly Cys Gly Leu Lys Leu Lys Val Asp Phe Arg Pro Ile  
 1570 1575 1580  
 Gly Leu Tyr Ala Gly Val Val Val Ala Pro Gly Leu Gly Ala Leu Pro  
 1585 1590 1595 1600  
 Asp Val Val Arg Phe Ala Gly Arg Leu Thr Glu Lys Asn Trp Gly Pro  
 1605 1610 1615  
 Gly Pro Glu Arg Ala Glu Gln Leu Arg Leu Ala Val Ser Asp Phe Leu  
 1620 1625 1630  
 Arg Lys Leu Thr Asn Val Ala Gln Met Cys Val Asp Val Val Ser Arg  
 1635 1640 1645  
 Val Tyr Gly Val Ser Pro Gly Leu Val His Asn Leu Ile Gly Met Leu  
 1650 1655 1660  
 Gln Ala Val Ala Asp Gly Lys Ala His Phe Thr Glu Ser Val Lys Pro  
 1665 1670 1675 1680  
 Val Leu Asp Leu Thr Asn Ser Ile Leu Cys Arg Val Glu  
 1685 1690

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 660 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Arg Pro Arg Pro Ile Leu Leu Leu Leu Leu Met Phe Leu Pro Met  
 1 5 10 15  
 Leu Pro Ala Pro Pro Pro Gly Gln Pro Ser Gly Arg Arg Arg Gly Arg  
 20 25 30  
 Arg Ser Gly Gly Ser Gly Gly Gly Phe Trp Gly Asp Arg Val Asp Ser  
 35 40 45  
 Gln Pro Phe Ala Ile Pro Tyr Ile His Pro Thr Asn Pro Phe Ala Pro  
 50 55 60  
 Asp Val Thr Ala Ala Ala Gly Ala Gly Pro Arg Val Arg Gln Pro Ala  
 65 70 75 80  
 Arg Pro Leu Gly Ser Ala Trp Arg Asp Gln Ala Gln Arg Pro Ala Val  
 85 90 95  
 Ala Ser Arg Arg Arg Pro Thr Thr Ala Gly Ala Ala Pro Leu Thr Ala  
 100 105 110  
 Val Ala Pro Ala His Asp Thr Pro Pro Val Pro Asp Val Asp Ser Arg  
 115 120 125  
 Gly Ala Ile Leu Arg Arg Gln Tyr Asn Leu Ser Thr Ser Pro Leu Thr  
 130 135 140



Ser Ser Val Ala Thr Gly Thr Asn Leu Val Leu Tyr Ala Ala Pro Leu  
 145 150 155 160  
 Ser Pro Leu Leu Pro Leu Gln Asp Gly Thr Asn Thr His Ile Met Ala  
 165 170 175  
 Thr Glu Ala Ser Asn Tyr Ala Gln Tyr Arg Val Ala Arg Ala Thr Ile  
 180 185 190  
 Arg Tyr Arg Pro Leu Val Pro Asn Ala Val Gly Gly Tyr Ala Ile Ser  
 195 200 205  
 Ile Ser Phe Trp Pro Gln Thr Thr Thr Thr Pro Thr Ser Val Asp Met  
 210 215 220  
 Asn Ser Ile Thr Ser Thr Asp Val Arg Ile Leu Val Gln Pro Gly Ile  
 225 230 235 240  
 Ala Ser Glu Leu Val Ile Pro Ser Glu Arg Leu His Tyr Arg Asn Gln  
 245 250 255  
 Gly Trp Arg Ser Val Glu Thr Ser Gly Val Ala Glu Glu Glu Ala Thr  
 260 265 270  
 Ser Gly Leu Val Met Leu Cys Ile His Gly Ser Leu Val Asn Ser Tyr  
 275 280 285  
 Thr Asn Thr Pro Tyr Thr Gly Ala Leu Gly Leu Leu Asp Phe Ala Leu  
 290 295 300  
 Glu Leu Glu Phe Arg Asn Leu Thr Pro Gly Asn Thr Asn Thr Arg Val  
 305 310 315 320  
 Ser Arg Tyr Ser Ser Thr Ala Arg His Arg Leu Arg Arg Gly Ala Asp  
 325 330 335  
 Gly Thr Ala Glu Leu Thr Thr Thr Ala Ala Thr Arg Phe Met Lys Asp  
 340 345 350  
 Leu Tyr Phe Thr Ser Thr Asn Gly Val Gly Glu Ile Gly Arg Gly Ile  
 355 360 365  
 Ala Leu Thr Leu Phe Asn Leu Ala Asp Thr Leu Leu Gly Gly Leu Pro  
 370 375 380  
 Thr Glu Leu Ile Ser Ser Ala Gly Gly Gln Leu Phe Tyr Ser Arg Pro  
 385 390 395 400  
 Val Val Ser Ala Asn Gly Glu Pro Thr Val Lys Leu Tyr Thr Ser Val  
 405 410 415  
 Glu Asn Ala Gln Gln Asp Lys Gly Ile Ala Ile Pro His Asp Ile Asp  
 420 425 430  
 Leu Gly Glu Ser Arg Val Val Ile Gln Asp Tyr Asp Asn Gln His Glu  
 435 440 445  
 Gln Asp Arg Pro Thr Pro Ser Pro Ala Pro Ser Arg Pro Phe Ser Val  
 450 455 460

Leu Arg Ala Asn Asp Val Leu Trp Leu Ser Leu Thr Ala Ala Glu Tyr  
 465 470 475 480  
 Asp Gln Ser Thr Tyr Gly Ser Ser Thr Gly Pro Val Tyr Val Ser Asp  
 485 490 495  
 Ser Val Thr Leu Val Asn Val Ala Thr Gly Ala Gln Ala Val Ala Arg  
 500 505 510  
 Ser Leu Asp Trp Thr Lys Val Thr Leu Asp Gly Arg Pro Leu Ser Thr  
 515 520 525  
 Ile Gln Gln Tyr Ser Lys Thr Phe Phe Val Leu Pro Leu Arg Gly Lys  
 530 535 540  
 Leu Ser Phe Trp Glu Ala Gly Thr Thr Lys Ala Gly Tyr Pro Tyr Asn  
 545 550 555 560  
 Tyr Asn Thr Thr Ala Ser Asp Gln Leu Leu Val Glu Asn Ala Ala Gly  
 565 570 575  
 His Arg Val Ala Ile Ser Thr Tyr Thr Thr Ser Leu Gly Ala Gly Pro  
 580 585 590  
 Val Ser Ile Ser Ala Val Ala Val Leu Ala Pro His Ser Ala Leu Ala  
 595 600 605  
 Leu Leu Glu Asp Thr Leu Asp Tyr Pro Ala Arg Ala His Thr Phe Asp  
 610 615 620  
 Asp Phe Cys Pro Glu Cys Arg Pro Leu Gly Leu Gln Gly Cys Ala Phe  
 625 630 635 640  
 Gln Ser Thr Val Ala Glu Leu Gln Arg Leu Lys Met Lys Val Gly Lys  
 645 650 655  
 Thr Arg Glu Leu  
 660

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 123 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Asn Asn Met Ser Phe Ala Ala Pro Met Gly Ser Arg Pro Cys Ala  
 1 5 10 15  
 Leu Gly Leu Phe Cys Cys Cys Ser Ser Cys Phe Cys Leu Cys Cys Pro  
 20 25 30  
 Arg His Arg Pro Val Ser Arg Leu Ala Ala Val Val Gly Gly Ala Ala  
 35 40 45  
 Ala Val Pro Ala Val Val Ser Gly Val Thr Gly Leu Ile Leu Ser Pro  
 50 55 60

Ser Gln Ser Pro Ile Phe Ile Gln Pro Thr Pro Ser Pro Pro Met Ser  
65 70 75 80  
Pro Leu Arg Pro Gly Leu Asp Leu Val Phe Ala Asn Pro Pro Asp His  
85 90 95  
Ser Ala Pro Leu Gly Val Thr Arg Pro Ser Ala Pro Pro Leu Pro His  
100 105 110  
Val Val Asp Leu Pro Gln Leu Gly Pro Arg Arg  
115 120

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7171 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Composite Mexico strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCCATGGAGG CCCACCAATT CATTAAGGCT CCTGGCATCA CTACTGCTAT TGAGCAAGCA	60
GCTCTAGCAG CGGCCAACTC CGCCCTTGCG AATGCTGTGG TGGTCCGGCC TTCCTTTCC	120
CATCAGCAGG TTGAGATCCT TATAAATCTC ATGCAACCTC GGCAGCTGGT GTTTCGTCCT	180
GAGGTTTTTTT GGAATCACCC GATTCAACGT GTTATACATA ATGAGCTTGA GCAGTATTGC	240
CGTGCTCGCT CGGGTCGCTG CCTTSAGATT GGAGCCCACC CACGCTCCAT TAATGATAAT	300
CCTAATGTCC TCCATCGCTG CTTTCTCCAC CCCGTCGGCC GGGATGTTCA GCGCTGGTAC	360
ACAGCCCCGA CTAGGGGACC TGCGGCGAAC TGTCGCCGCT CGGCACTTCG TGGTCTGCCA	420
CCAGCCGACC GCACTTACTG TTTTGATGGC TTTGCCGGCT GCGGTTTTGC CGCCGAGACT	480
GGTGTGGCTC TCTATTCTCT CCATGACTTG CAGCCGCTG ATGTTGCCGA GGCGATGGCT	540
CGCCAGGSCA TGAGCCGCTT TTATGCAGCT TTCACTTTC CTCCAGAGGT GCTCCTGCCT	600
CCTGGCACCT ACCGGACATC ATCTACTTG CTGATCCACG ATGGTAAGCG CGCGSTTGTC	660
ACTTATGAGG GTGACACTAG CGCCGGTTAC AATCATGATG TTGCCACCTT CGGCACATGG	720
ATCAGSACAA CTAAGGTGTG GGTGAACAC CTTTGGTGA TCGAGCGGGT GCGGGGTATT	780
GGCTGTCACT TTGTGTTGTT GATCACTGCG GCCCCGAGC CCTCCCCGAT GGCCTAGTT	840
CCTTACCCGC GTTCGACGGA GGTCTATGTC CGGTCTATCT TTGGGCCCCG CGGGTCCCCG	900

TGGGTGTTCC CGACCGGTTG TGGTGGAAG TGGATTTTC AGGCGGTCCC CAGGACATC	960
TGGGAGCGTC TCATGCTCTT TGGGCGAAG CTCGACGACC AGGCTTTTTG CTGCTCCAGG	1020
CTTATGAGT ACCTTCTGG CATTAGCTAT AAGGTAAGTG TGGGTGCGGT GGTGGTAAT	1080
GAAGGCTGGA ATGCGACCGA GGATGGGCTC ACTGCACTTA TTACGGCGGC TTACCTCACA	1140
ATATGTCATC AGGCTTATTT GGGGAGCGAG GCGATTTCTA AGGGCATGGG CCGGCTTGAG	1200
CTTGACATG CTCAGAAATT TATTTGAGC CTCTACAGCT GGTATTGGA GAATCAGGT	1260
CGTGATTACA TCGCAGCGG CAGCTGCG TTCTAGGCTC AGTGGCGGCG CTGCTTATCT	1320
GGCGGCTTCC AGCTGAGCC CCGCAGCTTA GTTTTGTATG AGTCAGTGGC TGTAGCTGC	1380
CGAACGACCA TCGGCGGAT CGCTGGAAGA TTTTGCTGT TTATGAAGTG GCTCGGTCAG	1440
GAGTGTCTT GTTTGGTCCA GCGCGCGAG GGGCTGGGG GCGACGAGG TCATGACAAT	1500
GAGGCTATG AAGGCTCTGA TGTTGATACT GCTGAGCCTG CCACCTAGA CATTACAGGC	1560
TCATACATCG TGGATGCTCG GTCTCTGCAA ACTGTCTATC AAGCTCTGGA CCTGCAAGT	1620
GACCTGGTAG CTGGCGGAGC CGAGTGTCT GCTACAGTTA CTGTTACTGA AACCTCTGGC	1680
CGTCTGGATT GCCAAACAAT GATCGGCAAT AAGACTTTTC TCACTACCTT TGTGATGGG	1740
GCACGCCTTG AGTTAACGG GCTGAGCAG CTTAACCTCT CTTTGGACAG CCAGCAGTGT	1800
AGTATGGCAG CCGGCCGTT TTGGCTCACC TATGCTGCG TAGATGGCGG GCTGGAAGTT	1860
CATTTTTCCA CCGCTGGCT CGAGAGCGGT GTTGTTTTCC CCCCTGGTAA TGCTCGGACT	1920
GCCCCGCCGA GTGAGGTCAC CGCTTCTGC TCAGCTCTTT ATAGGCACAA CCGCAGAGC	1980
CAGCGCCAGT CGGTTATGG TAGTTTGTG CTGCACCTG AAGGTTTGCT CCGCTGTTC	2040
CCGCCCTTTT CACCGGCGCA TGAGTGGCG TCTGCTAACC CATTTTGGCG CGAGAGCACG	2100
CTCTACACC GCACTTGGTC CACAATTACA GACACACCCT TAACTGTGG GCTAATTTCC	2160
GGTCATTTGG ATGCTGTCC CCACTCGGG GGGCCACCTG CTAAGCCAC AGGCCCTGCT	2220
GTAGGCTCGT CTGACTCTCC AGACCCTGAC CGCTACCTG ATGTTACAGA TGGCTCACGC	2280
CCCTCTGGGG CCCGTCCGGC TGGCCCCAAC CGGAATGGCG TTCCGAGCG CCGCTTACTA	2340
CACACCTACC CTGACGCGC TAAGATCTAT GTGGGTCCA TTTTGGAGTC TGAGTGCACC	2400
TGGTTGTCA ACGCATCTAA CGCGGGGAC CGGCTGTGTG GCGGCTTTG TCATGCTTTT	2460
TTTCAGCGTT ACCTGATTC GTTTGACGG ACCAAGTTTG TGATGCTGA TGGTCTTGCC	2520
GGTATACCC TTACAGCGG GGTGATCAT CATGCGGTGG CCGCGACTA TGATTGGAA	2580
CATAACCCCA AGAGGTGGA GGCTGCCTAC CGCGAGACTT GCGCGGCGG AGGCACTGCT	2640
GGTATCCAC TCTTAGCGG TGCCATTTAC CAGGTGCTG TTAGTTTGA TTTTGATGCC	2700
TGGGAGCGGA ACTACCGCG GTTTGACGAG CTTTACCTAA CAGAGCTGG GCTCGGTGG	2760

TTTBAATCCA ACCGCCCCGG TCAGCCCCAG TTBAACATAA CTGAGGATAC CGCCCCGTGG	2820
CTCAAGCTGG CCGTGGAGCT TGAATCGGG AGTBAAGTAG GCGCGGATG TGCGCGGTGT	2830
AAATCGAGC CTGGCGTTGT GCGGTATCAG TTTACAGCG GTGTCCCGG CTCTGCCAAG	2940
TCAAAGTCCG TGCAACAGGC GGATGTGGAT GTTGTGTGTG TGCCCACTCG CBAAGTTCCG	3000
AACGCTTGGC GCGCGCGGGG CTTTGGGCA TTCACTCGC AACTCGGC CGGTCTACT	3060
AGCGCGCTA GCGTTGTCT TGAATAGGC CCGTGGTTC CCGCACTT GGTGTTTTTA	3120
CATATGCAGC GTGTGTCTC TGTGCACTC GTTGGGAGC CCAATCAGAT CCGCGCGATA	3180
GATTTTBAAC AAGCGGTCT GATTCAAGCA ATACGCGCG AATTGGTTC GACTTCATGG	3240
TGGCATGTCA CCGCGGTGT CCGTCAAGAT GTGTGTGAT TAGTCCGTGG TGCTTACCT	3300
AAAATCCAGA CTACAAATAA GGTGTCTCT TCGTTTTT GTGGAGAGCC AGCTGTGGC	3360
CAGAAGTAG TGTCAACA GGTGTCTAAG GCGCGCGAC CCGCATCTAT AACGCTCCAT	3420
GAGGCGAGG GTGCACTTT TACCACTACA ACTATAATTG CAATGCAGA TGCCCGTGGC	3480
CTCATACAT CCGCGCGG TCACCTATA GTTGTCTCTA CTAGGCATAC TGAAAAATGT	3540
GTTATACTTG ACTCTCCCG CCGTGTGCT GAGGTGGTA TCTCAGATGC CATTGTTAAT	3600
AATTTCTTC TTTCCGGTGG CGAGGTGTGT CACCAAGAGC CATCGGTCT TCCGCGAGGC	3660
AACCTGACC GCAATGTTGA CGTGTGTGG CGGTTTCCAC CTTTATGCCA AATAAGCGCC	3720
TTCCATCAGC TTGCTGAGGA GCTGGGCCAC CGGCGCGGC CGGTGGCGGC TGTGCTACCT	3780
CGTGGCCTG AGCTTGAGCA GGGCCTTCTC TATCTGCCAC AGGAGCTAGC CTCCTGTGAC	3840
AGTGTGTGA CATTTGAGCT AACTGACATT GTGCACTGCC GCATGGCGGC CCGTAGCCAA	3900
AGGAAAGCTG TTTTGTCCAC GCTGGTAGGC CGGTATGGCA GACGCACAAG GCTTTATGAT	3960
GCGGGTCACA CCGATGTCCG CGCCTCCCTT GCGGCTTTA TTCCACTCT CGGGCGGGTT	4020
ACTGCCACCA CCGTGAAT CTTTGAAGCT GTAGAGGCGA TGGTGGAGAA GGGCCAAGAC	4080
GGTTCAGCG TCCCTGAGTT GGATTTGTG AGCGAGATG TCTCCCGCAT AACCTTTTC	4140
CAGAAGGATT GTAACAAGTT CACGACCGC GAGACAATTG CGCATGGCAA AGTCGGTCAG	4200
GGTATCTTC GCTGGAGTAA GACGTTTTGT GCGTGTGTG GCGCTGTGT CCGTGTGATT	4260
GAGAAGGCTA TTCTATCTT TTTACACAA GGTGTGTGT ACGGGATTC TTATACGAC	4320
TCAATATTCT CTGTGCGGT GGCTGGCGC AGCGATGCA TGGTGTGTA AAATGATTTT	4380
TCTGAGTTTG ACTGACTCA GAATACTTT TCGTAGGTC TTGATGCGC CATTATGAA	4440
GAGTGTGTA TGCCCGAGT GCTTGTAGG TTGTAGATG CCGTGGGTG GCGGTGGATC	4500
CTGCAGGCG CAAAAGATC TTTGAGAGG TTCTGGAAGA ACGATTCTG TGAGCGGGC	4560
AGCTTGCTCT GGAATACGT GTGGAAGAT GGAATCATG CCGATTGTA TGAGTTCCGG	4620

GACCTCCAGG TTGCGGCGTT CAAGGGGAG GACTCGGTGG TGGTGTGTAG TGAATATGGC	4680
CAGAGCGGAG CGGCTGGTTC GCTTATAGCA GGTGTGTGT TGAAGTTGAA GGTGAGTTC	4740
CGGCGGATG GGTGTGTATG CGGGGTTGTG GTGGCGGGG GGTGCGGGG CTTACCGGAT	4800
GTGGTTGGAT TGCGCGGAG GCTTTGGAG AAGAACTGG GGGCTGATCC GGAGCGGGCA	4860
GAGTAGGTCC GCGTGGCGGT GCAAGGATTC CTGGTAGGT TAAAGAAAT GTGGCGAGAT	4920
TGTGTTGAGG TGGTGTGTAG AGTTTACGG GTTTCCCGGG GTCTGGTTCA TAACTGATA	4980
GGATGCTCC AAGTATTGG TGATGTGAAG GGGCATTTTA CAGAGTGTGT TAAGCTATA	5040
TTTGAGTTTA CAGACTCAAT TATGACCGG TGTGAATGAA TAAATGTGT TTTGTCGGC	5100
CGATGCTTC GCGACGATG GCGTAGGCG TGTGTTGGTG TTGTTGCTT TGTGTTGCG	5160
TATGTTGCGG GTGGGACCGA CGGTCAGCG GTGTGGGGG CGTGTGGGG GGCGCAGCGG	5220
CGGTACCGG GGTGTTTGT GGGGTGACCG GGTGATTCT CAGCGCTTG CAATCCCGTA	5280
TATTCATCCA ACCAAGCGGT TTGCGCGAGA GTTGCGCGT GCGTCGGGT CTGGAGCTG	5340
CGTTCGGGAA CGAGCGGGG CAGTGGGTC CAGTGGGGA GATCAGGCG AGCGCGGCTC	5400
CGCTGCGTCC GGTGCGGAG CTGGCAGAG CGGGGCTGG GCGCTGACCG CTGTGGGCGC	5460
TGCGCATGAC AGCTACCGG TCGCGGAGT TGATTCTGG GGTGCAATC TACGCGGCA	5520
GTATAATTT TCTACTTCAC CCCTGACATC CTGTGTGGC TCTGGCACTA ATTTAGTCCT	5580
GTATGAGCG CGCTTAATC CGCTCTGCC GCTGCAGGAC GGTACTAATA CTCACATTAT	5640
GGCGACAGG GCTCCAATT ATGCACAGTA CGGGGTTGCC CGCGCTACTA TCGGTTACCG	5700
GCGCTAGTG CCTAATGCAG TTGGAGGCTA TGCTATATC ATTTCTTTCT GGCCTCAAAC	5760
AACCAACAAC CCTACATCTG TTGACATGAA TTCCATTACT TCCACTGATG TCAGGATTCT	5820
TGTTCAACCT GGCATAGCAT CTGAATTGGT CATCCCAAGC GAGCGCTTC ACTACCGCAA	5880
TCAAGGTTGG CGCTCGGTTG AGACATCTGG TGTGCTGAG GAGGAAGCCA CCTCCGGTCT	5940
TGTGATGTTA TGCATACATG GCTCTCCAGT TAACTCCTAT ACCAATACCG CTTATACCGG	6000
TGCGCTTGGC TTAAGGACT TTGCCTTAGA GCTTGAGTTT CGCAATCTCA CCACCTGTAA	6060
CAGCAATACA CGTGTGTGCC GTTACTCCAG CACTGCTGT CACTCGGCG GAGGGGGCGA	6120
CGGCACTGG GAGGTGACCA CAACTGAGC CAGCAGGTTT ATGAAAGAT TCGACTTTAC	6180
CGGCTTAAT GGGTAGGTG AAGTGGGCG CGGGATAGCT CTAACATTAC TTAACCTTGC	6240
TGACAGGCTC CTGGGCGGG TCGGACAGA ATTAATTTG TCGGTGGG GGCAACTGTT	6300
TTATTCCCGC CGGTTGTCT CAGCCAATGG CGAGCAACG GTGAAGCTCT ATACATCAGT	6360
GGAGAAATGT CAGCAGGATA AGGTTGTTC TATCCCGAC GATATGATC TTGGTATTC	6420
GCGTGTGTC ATTAGGATT ATGACAACCA GCATGAGCAG GATCGGCGA CCGCTCGCG	6480

TSGGCCATCT CGGCCTTTTT CTGTTCTGCG AGCAAAATGAT GTACTTTGGC TGTCCCTCAC	6540
TGAGCCGAG TATGACCACT CCACTTACCG GTGCTCAACT GGGCCGGTTT ATATCTCGGA	6600
CAGCGTCACT TTGCTGAATG TTGCGACTGG CGCGCAGGCC STAGCCCGAT CGCTTCACTG	6660
GTCCAAAGTC ACCCTCGACG GCGCCCCCT CCGACTGTT GAGCAATATT CCAAGACATT	6720
CTTGTGCTC CCGCTTCGTG GCAAGCTCTC CTTTTGGGAG GCCGGCACAA CAAAAGCAGG	6780
TTATGCTTAT AATTATAATA CTACTGCTAG TGACCAAGATT CTGATTGAAA ATGCTGCCGG	6840
CGATCGGCTC GCGATTTCAA CCGTAACAC CAGGCTTGGG GCGGGTGGG TCGCATTTTC	6900
TGCGCGCGCG GTTTTGGCTC CAGGCTCGCG CCGGCTCTG CTGGAGGATA CTTTGTATTA	6960
TGCGCGCGCG GCGCACACAT TTGATGACTT CTGCGCTGAA TGCGCGGCTT TAGGCTTCCA	7020
GGGTTGTGCT TTCCAGTCAA CTGTGCTGA GCTCAGGCG CTTAAAGTTA AGGTGGGTAA	7080
AACTCGGGAG TTGTAGTTTA TTTGGCTGTG CCGACCTACT TATATCTGCT GATTTCCTTT	7140
ATTTCTTTTT TCTCGSTCCC GCGCTCCTG A	7171

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1575 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: T: Mexican strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTTGGGTGAG GTGGGTATCT CAGATGCCAT TGTTAATAAT TTCTTCTTT CGGGTGGCGA	60
GGTGGTCAAG CAGAGACCAT CGGTCAATCC GCGAGGCAAC CCGACCGCA ATGTTGACGT	120
GCTGGGGCG TTTCAGCTT CATGCCAAAT AAGCGCTTC CATCAGCTTG CTGAGGAGCT	180
GGGCAACCG CGCGCGCGG TGGCGGCTGT GCTAGCTCC TCGCTGAGC TTGAGCAGG	240
CCTTCTTAT CTGCCACAGG AGCTAGCTC CTGTGACAGT GTTGTGACAT TTGAGCTAAC	300
TGACATTGTC CACTGCGCA TGGCGCGCG TAGCCAAAGG AAAGCTGTT TGTCACGCT	360
GGTAGGCGG TATGCCAGAC GCACAAGCT TTATGATGCG GGTACACCG ATGTCCGCG	420
CTCCCTTGCG CGCTTTATTC CCACTCTGG GCGGTTACT GCCACCACCT GTGAACTCTT	480
TGAGCTTGTA GAGGCGATGG TGGAGAAGG CCAAGACGCT TCAGCCGTC TCGACTTGA	540

TTTGTGAGG CGAGATGTGT GTGGGATAAG GTTTTTCAG AAGGATTGTA AUAAGTTCAC	600
GAGGGGGGAG ACAATTGGGG ATGGGAAAGT GGGTCAGGGT ATCTTGGGCT GGAGTAAGAG	660
CTTTTGTGGG GTGTTTGGGG CCGGTTCCG TGGATTGAG AAGGCTATTG TATGCTTTT	720
ACCACAAGCT GTGTTCTAGG GGGATGCTTA TGACGACTCA GTATTCTCTG CTGCGGTGGC	780
TGGGGCCAGG CATGCCATGG TGTTCGAAAA TGATTTTTCT GAGTTTGACT CGACTCAGAA	840
TAACTTTTCC CTAGGTCTTG AGTGGGCCAT TATGGAAGAG TGTGGTATGC CCCAGTGGCT	900
TGTGAGGTTG TACCATGCGG TCGGGTGGG GTGGATCTG CAGGCCCCAA AAGAGTCTTT	960
GAGAGGTTG TGGAAGAAGC ATTCTGGTGA GCGGGGACG TTGCTCTGGA ATACGGTGTG	1020
GAACTGGCA ATCATTGCGG ATTGCTATGA GTTGGGGAC CTCAGGTTG CCGCTTCAA	1080
GGGGGAGGAC TCGGTCGTG TCTGTAGTGA ATACGGGAG AGCCGAGGGG CCGTTCTCT	1140
TATAGCAGG GTTGGTTTGA AGTTGAAGGC TGACTTGGG CCGATTGGGC TGTATGCGG	1200
GGTTGTCTG GCGGGGGGG TCGGGGCGCT ACCCGATGTC GTTCGATTGG CCGGAGGGCT	1260
TTGGGAGAAG AACTGGGGG CTGATCGGA GCGGGCAGAG CAGCTCGGC TCGCGGTGCA	1320
GGATTTCTC CTAGGTTAA CGAATGTGGC CCAGATTTGT GTTGAGGTGG TGTCTAGAGT	1380
TTACGGGGTT TCGCGGGTC TGGTTCATAA CCTGATAGGC ATGCTCCAGA CTATTGGTGA	1440
TGGTAAGGCG CATTTTACAG AGTCTGTAA GCCTATACTT GACCTTACAC ACTCAATTAT	1500
GCACCGGTCT GAATGAATAA CATGTGTTT GTTGCGCCCA TGGGTTGCGC ACCATGCGCC	1560
CTAGGCCTCT TTTGC	1575

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 874 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Tashkent strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGGGGCGCGT ACAGGTGACA ACCTGTGAGT TGTAGAGCT AGTGGAGGCG ATGGTCGAGA	60
AAGGGCAGGA TGGGTCGCGG GTGCTTGGC TGGATCTCTG CAACCGTGAC GTGTCCAGGA	120
TCACCTTTTT CCAGAAAGAT TGCAATAAGT TCACACGGG AGAGACCATC GCGCATGGTA	180



AAATGGGCGA GGGCATTTCG GCGTGGASTA AGACCTTTTG TGCCCTTTTC GGCCCTGGT	240
TCCGTGCTAT TGAGAAGGCT ATTCTGGCCC TGCTCCCTCA GGGTGTGTTT TATGGGGATG	300
CCTTTGATGA CACCTGCTTC TGGGCGGCTG TGGCGCGACC AAAGGCGTCC ATGGTGTGTTG	360
AGAATGACTT TTCTGAGTTT GACTCCACCC AGAATAATTT TTCCCTGGGC CTAGAGTGTG	420
CTATTATGGA GAAGTGTGGG ATGDCBAAGT GGCTCATCCG CTTGTACCAC CTTATAAGGT	480
CTSCGTGGAT DCTGAGGCG CCGAAGGAGT CCGTSCGAGG GTGTTGGAAG AAACACTCCG	540
GTSAGCCCGG CACTCTTCTA TGAATACTG TCTGGAACAT GGCGTTATC ACCCATTTGT	600
AGGATTTCCG CGATTTCAG GTGGCTGCGT TAAAGGTGA TGATTCGATA GTGCTTTGCA	660
GTGASTACCG TCAGAGTCCA GGGGCTGCTG TCGTGAATGC TGGTGTGGC TAAAGCTGA	720
AGGTGGGTTT CCGTCCGATT GGTGTTGATG CAGGTGTTGT GTGAGGCCCC GGCTTTGGCG	780
CGCTTCCCGA CGTCTGCGC TTGTCCGCGC GGCTTACTGA GAAGAATTGG GGCCCTGGCC	840
CTGAGCGGGC GGAGCAGCTC CGCCTTGCTG TGCG	874

(1) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 449 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone 406.4-2 cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..100

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

C GCC AAC CAG CCC GGC CAC TTG GCT CCA CTT GGC GAG ATC AGG CCC	46
Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro	
1 5 10 15	
AGC GGC CTT CCG CTG CCT GCG CTC GCC GAG CTG CCA CAG CCG GGG CTG	94
Ser Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu	
20 25 30	
CGG CCG TACCGGTGT GCGCGCTGCC CATGACACCT CACCGTCCC GGACGTTGAT	150
Arg Arg	
TCTCGCGGTG CAATTCTAGG CCGCCAGTAT AATTTGTCTA CTTACCCCT GACATCCTCT	210

GTGBCCTGTG GCACTAATTT AGTCCTGTAT GCAGCCCCCC TTAATCCGCG TCTGCCGCTG 270  
 CAGBACGGTA CTAATACTCA CATTATGGCC ACAGAGGGCT CCAATTATGC ACASTACGG 330  
 GTTCCCGCGG CTACTATCGG TTACCGGGCC CTAGTGCCTA ATGCAGTTGG AGGCTATGCT 390  
 ATATCCATTT CTTTCTGGCC TCAAACAACC ACAACCCCTA CATCTGTIGA CATGAATTC 449

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro Ser  
 1 5 10 15  
 Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu Arg  
 20 25 30  
 Arg

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 130 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: Clone 406.3-2

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 5..130

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGAT ACT TTT GAT TAT CCG GGG CGG GCG CAC ACA TTT GAT GAC TTC TGC 49  
 Thr Phe Asp Tyr Pro Gly Arg Ala His Thr Phe Asp Asp Phe Cys  
 1 5 10 15  
 GGT GAA TGC CGC GCT TTA GGC CTC CAG GGT TGT GCT TTC CAG TCA ACT 97  
 Pro Glu Cys Arg Ala Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr  
 20 25 30  
 GTC GCT GAG CTC CAG CGC CTT AAA GTT AAG GTT 130

Val Ala Glu Leu Gln Arg Leu Lys Val Lys Val  
 35 40

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 42 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Thr Phe Asp Tyr Pro Gly Arg Ala His Thr Phe Asp Asp Phe Cys Pro  
 1 5 10 15  
 Glu Cys Arg Ala Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Val  
 20 25 30  
 Ala Glu Leu Gln Arg Leu Lys Val Lys Val  
 35 40

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 406.4-2 epitope - Mexican strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ala Asn Gln Pro Gly His Leu Ala Pro Leu Gly Glu Ile Arg Pro Ser  
 1 5 10 15  
 Ala Pro Pro Leu Pro Pro Val Ala Asp Leu Pro Gln Pro Gly Leu Arg  
 20 25 30  
 Arg

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (vi) ORIGINAL SOURCE:  
       (C) INDIVIDUAL ISOLATE: 406.4-2 epitope - Burma strain  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ala	Asn	Pro	Pro	Asp	His	Ser	Ala	Pro	Leu	Gly	Val	Thr	Arg	Pro	Ser
1				5					10					15	
Ala	Pro	Pro	Leu	Pro	His	Val	Val	Asp	Leu	Pro	Gln	Leu	Gly	Pro	Arg
			20					25					30		

Arg

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 42 amino acids  
       (B) TYPE: amino acid  
       (C) STRANDEDNESS: single  
       (D) TOPOLOGY: unknown  
 (ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO  
 (vi) ORIGINAL SOURCE:  
       (C) INDIVIDUAL ISOLATE: 406.3-2 epitope - Mexican strain  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Thr	Phe	Asp	Tyr	Pro	Gly	Arg	Ala	His	Thr	Phe	Asp	Asp	Phe	Cys	Pro
1				5					10					15	
Glu	Cys	Arg	Ala	Leu	Gly	Leu	Gln	Gly	Cys	Ala	Phe	Gln	Ser	Thr	Val
			20					25					30		
Ala	Glu	Leu	Gln	Arg	Leu	Lys	Val	Lys	Val						
		35					40								

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 42 amino acids  
       (B) TYPE: amino acid  
       (C) STRANDEDNESS: single  
       (D) TOPOLOGY: unknown  
 (ii) MOLECULE TYPE: peptide  
 (iii) HYPOTHETICAL: NO  
 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(C) INDIVIDUAL ISOLATE: 406.3-2 epitope - Burma strain

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Thr Leu Asp Tyr Pro Ala Arg Ala His Thr Phe Asp Asp Phe Cys Pro  
1 5 10 15

Glu Cys Arg Pro Leu Gly Leu Gln Gly Cys Ala Phe Gln Ser Thr Val  
20 25 30

Ala Glu Leu Gln Arg Leu Lys Met Lys Val  
35 40